

RECEIVED #218



1600

ENTERED

## RAW SEQUENCE LISTING

DATE: 04/21/2003

PATENT APPLICATION: US/09/988,292A

TIME: 15:30:56

Input Set : A:\pfl60d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

## SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: Yu, Guo-Liang

7 Rosen, Craig

9 (ii) TITLE OF INVENTION: Colon Specific Genes and Proteins

11 (iii) NUMBER OF SEQUENCES: 24

13 (iv) CORRESPONDENCE ADDRESS:

14 (A) ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,

15 Stewart & Olstein

16 (B) STREET: 6 Becker Farm Road

17 (C) CITY: Roseland

18 (D) STATE: NJ

19 (E) COUNTRY: USA

20 (F) ZIP: 07068-1739

22 (v) COMPUTER READABLE FORM:

23 (A) MEDIUM TYPE: Floppy disk

24 (B) COMPUTER: IBM PC compatible

25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

26 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

28 (vi) CURRENT APPLICATION DATA:

C--> 29 (A) APPLICATION NUMBER: US/09/988,292A

C--> 30 (B) FILING DATE: 19-Nov-2001

31 (C) CLASSIFICATION:

33 (viii) ATTORNEY/AGENT INFORMATION:

34 (A) NAME: Ferraro, Gregory D.

35 (B) REGISTRATION NUMBER: 36,134

36 (C) REFERENCE/DOCKET NUMBER: 325800-435

38 (ix) TELECOMMUNICATION INFORMATION:

39 (A) TELEPHONE: 201-994-1700

40 (B) TELEFAX: 201-994-1744

43 (2) INFORMATION FOR SEQ ID NO: 1:

45 (i) SEQUENCE CHARACTERISTICS:

46 (A) LENGTH: 638 base pairs

47 (B) TYPE: nucleic acid

48 (C) STRANDEDNESS: single

49 (D) TOPOLOGY: linear

51 (ii) MOLECULE TYPE: cDNA

54 (ix) FEATURE:

55 (A) NAME/KEY: mat\_peptide

56 (B) LOCATION: 1..501

58 (ix) FEATURE:

59 (A) NAME/KEY: CDS

60 (B) LOCATION: 1..501

## RAW SEQUENCE LISTING

DATE: 04/21/2003

PATENT APPLICATION: US/09/988,292A

TIME: 15:30:56

Input Set : A:\pfl60d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

```

63      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
64 GCC AGG CAG CTG GCT GCC SAC CAG GCC GTG TAT GTG AAG GTC AAG GCT      48
W--> 65 Ala Arg Gln Leu Ala Ala Xaa Gln Ala Val Tyr Val Lys Val Lys Ala
66   1           5           10           15
68 GAA GCC CGG GAA CTG CTG GGC CAC CCG TGG TCT CTG TGT CCT GTC TGT      96
69 Glu Ala Arg Glu Leu Leu Gly His Pro Trp Ser Leu Cys Pro Val Cys
70           20           25           30
72 GGG TGC CAA CTC ACC ACC TTT GAT GGG GCC CGT GGT GCC ACC ACT CTC      144
73 Gly Cys Gln Leu Thr Thr Phe Asp Gly Ala Arg Gly Ala Thr Thr Leu
74           35           40           45
76 CTG GTG TCT ATG AAG CTC TCT TCC CGC TGC CCA GGA CTA CAG AAT ACC      192
77 Leu Val Ser Met Lys Leu Ser Ser Arg Cys Pro Gly Leu Gln Asn Thr
78   50           55           60
80 ATC CCC TGG TAC CGT GTA GTT GCC GAA GTC CAG ATC TGC CAT GGC AAA      240
81 Ile Pro Trp Tyr Arg Val Val Ala Glu Val Gln Ile Cys His Gly Lys
82 65           70           75           80
84 ACG GAG GCT GTG GGC CAG GTC CAC ATC TTC TTC CAG GAT GGG ATG GTG      288
85 Thr Glu Ala Val Gly Gln Val His Ile Phe Phe Gln Asp Gly Met Val
86           85           90           95
88 ACG TTG ACT CCA AAC AAG GGT GTG TGG GTG AAT GGT CTC CGA GTG GAT      336
89 Thr Leu Thr Pro Asn Lys Gly Val Trp Val Asn Gly Leu Arg Val Asp
90   100          105          110
92 CTC CCA GCT GAG AAG TTA GCA TCT GTG TCC GTG AGT CGT ACA CCT GAT      384
93 Leu Pro Ala Glu Lys Leu Ala Ser Val Ser Val Ser Arg Thr Pro Asp
94   115          120          125
96 GGC TCC CTG CTA GTC CGC CAG AAG GCA GGG GTC CAG GTG TGG CTT GGA      432
97 Gly Ser Leu Leu Val Arg Gln Lys Ala Gly Val Gln Val Trp Leu Gly
98   130          135          140
100 GCC AAT GGG AAG GTG GCT GTG ATT GTG AGC AAT GAC CAT GCT GGG AAA      480
101 Ala Asn Gly Lys Val Ala Val Ile Val Ser Asn Asp His Ala Gly Lys
102 145          150          155          160
104 CTG TGT GGG GGC CTK TGG AAA ATTTGACGGG GGACCAGACC AATGATTGGG      531
105 Leu Cys Gly Gly Xaa Trp Lys
106           165
108 ATGATTCCCA GGAGAAGCCA GCGATTGGGG AAWTGGAGAG CGCAGGGACT TTCTYCCMCA      591
110 TGTTAATGGG CTTGWTCCAG TTCATCCCAC CAGGAACGAA GGATTTT      638
113 (2) INFORMATION FOR SEQ ID NO: 2:
115      (i) SEQUENCE CHARACTERISTICS:
116          (A) LENGTH: 167 amino acids
117          (B) TYPE: amino acid
118          (D) TOPOLOGY: linear
120      (ii) MOLECULE TYPE: protein
122      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
W--> 124 Ala Arg Gln Leu Ala Ala Xaa Gln Ala Val Tyr Val Lys Val Lys Ala
125   1           5           10           15
127 Glu Ala Arg Glu Leu Leu Gly His Pro Trp Ser Leu Cys Pro Val Cys
128           20           25           30
130 Gly Cys Gln Leu Thr Thr Phe Asp Gly Ala Arg Gly Ala Thr Thr Leu
131           35           40           45

```

## RAW SEQUENCE LISTING

DATE: 04/21/2003

PATENT APPLICATION: US/09/988,292A

TIME: 15:30:57

Input Set : A:\pfl60d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

```

133 Leu Val Ser Met Lys Leu Ser Ser Arg Cys Pro Gly Leu Gln Asn Thr
134      50                      55                      60
136 Ile Pro Trp Tyr Arg Val Val Ala Glu Val Gln Ile Cys His Gly Lys
137 65                      70                      75                      80
139 Thr Glu Ala Val Gly Gln Val His Ile Phe Phe Gln Asp Gly Met Val
140      85                      90                      95
142 Thr Leu Thr Pro Asn Lys Gly Val Trp Val Asn Gly Leu Arg Val Asp
143      100                     105                     110
145 Leu Pro Ala Glu Lys Leu Ala Ser Val Ser Val Ser Arg Thr Pro Asp
146      115                     120                     125
148 Gly Ser Leu Leu Val Arg Gln Lys Ala Gly Val Gln Val Trp Leu Gly
149      130                     135                     140
151 Ala Asn Gly Lys Val Ala Val Ile Val Ser Asn Asp His Ala Gly Lys
152 145                      150                      155                      160
154 Leu Cys Gly Gly Xaa Trp Lys
155      165
157 (2) INFORMATION FOR SEQ ID NO: 3:
159     (i) SEQUENCE CHARACTERISTICS:
160         (A) LENGTH: 874 base pairs
161         (B) TYPE: nucleic acid
162         (C) STRANDEDNESS: single
163         (D) TOPOLOGY: linear
165     (ii) MOLECULE TYPE: cDNA
168     (ix) FEATURE:
169         (A) NAME/KEY: CDS
170         (B) LOCATION: 1..705
172     (ix) FEATURE:
173         (A) NAME/KEY: mat_peptide
174         (B) LOCATION: 1..705
177     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
179 CAG GAC TGC GTG TGC ACG GAC AAG GTG GAC AAC AAC ACC CTG CTC AAC      48
180 Gln Asp Cys Val Cys Thr Asp Lys Val Asp Asn Asn Thr Leu Leu Asn
181 1      5      10      15
183 GTC ATC GCC TGC ACC CAC GTG CCC TGC AAC ACC TCC TGC AGC CCT GGG      96
184 Val Ile Ala Cys Thr His Val Pro Cys Asn Thr Ser Cys Ser Pro Gly
185      20      25      30
187 TTC GAA CTC ATG GAG GCC CCC GGG GAG TGC TGT AAG AAG TGT GAA CAG      144
188 Phe Glu Leu Met Glu Ala Pro Gly Glu Cys Cys Lys Lys Cys Glu Gln
189      35      40      45
190 ACG CAC TGT ATC ATC AAA CGG CCC GAC AAC CAG CAC GTC ATC CTG AAG      192
191 Thr His Cys Ile Ile Lys Arg Pro Asp Asn Gln His Val Ile Leu Lys
192      50      55      60
194 CCC GGG GAC TTC AAG AGC GAC CCG AAG AAC AAC TGC ACA TTC TTC AGC      240
195 Pro Gly Asp Phe Lys Ser Asp Pro Lys Asn Asn Cys Thr Phe Phe Ser
196 65      70      75      80
198 TGC GTG AAG ATC CAC AAC CAG CTC ATC TCG TCC GTT TCC AAC ATC ACC      288
199 Cys Val Lys Ile His Asn Gln Leu Ile Ser Ser Val Ser Asn Ile Thr
200      85      90      95
202 TGC CCC AAC TTT GAT GCC AGC ATT TGC ATC CCG GGC TCC ATC ACA TTC      336

```

## RAW SEQUENCE LISTING

DATE: 04/21/2003

PATENT APPLICATION: US/09/988,292A

TIME: 15:30:57

Input Set : A:\pf160d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

```

203 Cys Pro Asn Phe Asp Ala Ser Ile Cys Ile Pro Gly Ser Ile Thr Phe
204      100      105      110
206 ATG CCC AAT GGA TGC TGC AAG ACC TGC ACC CCT CGC AAT GAG ACC AGG      384
207 Met Pro Asn Gly Cys Cys Lys Thr Cys Thr Pro Arg Asn Glu Thr Arg
208      115      120      125
210 GTG CCC TGC TCC ACC GTC CCC GTC ACC ACG GAG GTT TCG TAC GCC GGC      432
211 Val Pro Cys Ser Thr Val Pro Val Thr Thr Glu Val Ser Tyr Ala Gly
212      130      135      140
214 TGC ACC AAG ACC GTC CTC ATG AAT CAT TGC TCC GGG TCC TGC GGG ACA      480
215 Cys Thr Lys Thr Val Leu Met Asn His Cys Ser Gly Ser Cys Gly Thr
216 145      150      155      160
218 TTT GTC ATG TAC TCG GCC AAG GCC CAG GCC CTG GAC CAC AGC TGC TCC      528
219 Phe Val Met Tyr Ser Ala Lys Ala Gln Ala Leu Asp His Ser Cys Ser
220      165      170      175
222 TGC TGC AAA GAG GAG AAA ACC AGC CAG CGT GAG GTG GTC CTG AGC TGC      576
223 Cys Cys Lys Glu Glu Lys Thr Ser Gln Arg Glu Val Val Leu Ser Cys
224      180      185      190
226 CCC AAT GGC GGC TCG CTG ACA CAC ACC TAC ACC CAC ATC GAG AGC TGC      624
227 Pro Asn Gly Gly Ser Leu Thr His Thr Tyr Thr His Ile Glu Ser Cys
228      195      200      205
230 CAG TGC CAG GAC ACC GTC TGC GGG CTC CCC ACC GGC ACC TCC CGC CGG      672
231 Gln Cys Gln Asp Thr Val Cys Gly Leu Pro Thr Gly Thr Ser Arg Arg
232      210      215      220
234 GCC CGG CGT TCC CCT AGG CAT CTG GGG AGC GGG TGAGCGGGGT GGGCACAGCC      725
235 Ala Arg Arg Ser Pro Arg His Leu Gly Ser Gly
236 225      230      235
238 CCTTCACTGC CCTCGACAGC TTTACCTCCC CCGGACCCTC TGAGCCTCCT AAGCTCGGCT      785
240 TCCTCTCTTC AGATATTTAT TGTCTGAGTT TTTGTTCAGT CCTTGCTTTC CAATAATAAA      845
242 CTCAGGGGGA CATGCAAAAA AAAAAAAAAA      874
245 (2) INFORMATION FOR SEQ ID NO: 4:
247 (i) SEQUENCE CHARACTERISTICS:
248 (A) LENGTH: 235 amino acids
249 (B) TYPE: amino acid
250 (D) TOPOLOGY: linear
252 (ii) MOLECULE TYPE: protein
253 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
255 Gln Asp Cys Val Cys Thr Asp Lys Val Asp Asn Asn Thr Leu Leu Asn
256 1      5      10      15
258 Val Ile Ala Cys Thr His Val Pro Cys Asn Thr Ser Cys Ser Pro Gly
259      20      25      30
261 Phe Glu Leu Met Glu Ala Pro Gly Glu Cys Cys Lys Lys Cys Glu Gln
262      35      40      45
264 Thr His Cys Ile Ile Lys Arg Pro Asp Asn Gln His Val Ile Leu Lys
265      50      55      60
267 Pro Gly Asp Phe Lys Ser Asp Pro Lys Asn Asn Cys Thr Phe Phe Ser
268 65      70      75      80
270 Cys Val Lys Ile His Asn Gln Leu Ile Ser Ser Val Ser Asn Ile Thr
271      85      90      95
273 Cys Pro Asn Phe Asp Ala Ser Ile Cys Ile Pro Gly Ser Ile Thr Phe

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292A

DATE: 04/21/2003

TIME: 15:30:57

Input Set : A:\pfl160d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

```

274          100          105          110
276 Met Pro Asn Gly Cys Cys Lys Thr Cys Thr Pro Arg Asn Glu Thr Arg
277          115          120          125
279 Val Pro Cys Ser Thr Val Pro Val Thr Thr Glu Val Ser Tyr Ala Gly
280          130          135          140
282 Cys Thr Lys Thr Val Leu Met Asn His Cys Ser Gly Ser Cys Gly Thr
283 145          150          155          160
285 Phe Val Met Tyr Ser Ala Lys Ala Gln Ala Leu Asp His Ser Cys Ser
286          165          170          175
288 Cys Cys Lys Glu Lys Thr Ser Gln Arg Glu Val Val Leu Ser Cys
289          180          185          190
291 Pro Asn Gly Gly Ser Leu Thr His Thr Tyr Thr His Ile Glu Ser Cys
292          195          200          205
294 Gln Cys Gln Asp Thr Val Cys Gly Leu Pro Thr Gly Thr Ser Arg Arg
295          210          215          220
297 Ala Arg Arg Ser Pro Arg His Leu Gly Ser Gly
298 225          230          235
300 (2) INFORMATION FOR SEQ ID NO: 5:
302     (i) SEQUENCE CHARACTERISTICS:
303         (A) LENGTH: 1209 base pairs
304         (B) TYPE: nucleic acid
305         (C) STRANDEDNESS: single
306         (D) TOPOLOGY: linear
308     (ii) MOLECULE TYPE: cDNA
310     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
315 ATTGGTGCTA CCTGGCTCTC CTGTCTCTGC AGCTCTACAG GTGAGGCCCA GCAGAGGGAG      60
316 TAGGGCTCGC CATGTTTCTG GTGAGCCAAT TTGGCTGATC TTGGGTGTCT GAACAGCTAT      120
318 TGGGTCCACC CCAGTCCCTT TCAGCTGCTG CTTAATGCCC TGCTCTCTCC CTGGCCCACC      180
320 TTATAGAGAG CCCAAAGAGC TCCTGTAAGA GGGAGAACTC TATCTGTGGT TTATAATCTT      240
322 GCACGAGGCA CCAGAAGTCT CCCTGGGTCT TGTGAATGAA CTACATTTAT CCCCTTTCCT      300
324 GCCCCAACCA CAAACTCTTT CCTTCAAAGA GGGCCTGCCT GGTTCCTTCC ACCCAACTGC      360
326 ACCATGAGAT CGGTCCAAGA GTCCATTCCC CAGGTGGGAG CCAACTGTCA GGGAGGTCTT      420
328 TCCCACCAAA CATCTTTCAG TTGCTGGGAG GTGACCATAG GGCTCTGCTT TTAAAGATAT      480
330 GGCTGCTTCA AAGGCCAGAG TCACAGGAAG GACTTCTTCC AGGGAGATTA GTGGTGATGG      540
332 AGAGGAGAGT TAAAAATGACC TCATGTCTCT CTTGTCCACG GTTTTGTGTA GTTTTCACTC      600
334 TTCTAATGCA AGGGTCTCAC ACTGTGAACC ACTTAGGATG TGATCACTTT CAGGTGGCCA      660
336 GGAATGTTGA ATGTCTTTGG CTCAGTTCAT CTAAAAAAGA TATCTATTTG AAAGTTCTCA      720
338 GAGTTGTACA TATGTTTCAC AGTACAGGAT CTGTACATAA AAGTTTCTTT CCTAAACCAT      780
340 TCACCAAGAG CCAATATCTA GGCATTTTCT CGGTAGCACA AATTTTCTNA TTGCTTAGAA      840
342 AATTGTCTC CTGTTCTTTT CTGTCTGNAG ACTTAAGTGA GTTAGGTCTT TAAGGAAAGC      900
344 AACGCTCCTC TGAAATGCTT GTCTTTTTTC TGTGCCGAA ATAGCTGGTC CTTTTTCGGG      960
346 AGTTAGATGT ATAGAGTGTT TGTATGTAAA CATTTCTTGT AGGCATCACC ATGAACANAG      1020
348 ATATATTTTC TATTTANTTA NTATATGTGC ACTTCAAGAA GTCACGTGTA GAGAAATAAA      1080
350 GAATGTCTT AAATGTCATG ATTGGAGATG TCCTTTGCAT TGCTTGGAAG GGGTGTACCT      1140
352 AGAGCCAAGG AAATTGGCTC TGGTTTGGA AAATTTTGCT GTTATTATAG TAAACATACA      1200
354 AAGGATGTC
356 (2) INFORMATION FOR SEQ ID NO: 6:
358     (i) SEQUENCE CHARACTERISTICS:
359         (A) LENGTH: 548 base pairs

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/988,292A

DATE: 04/21/2003  
TIME: 15:30:58

Input Set : A:\pfl60d2seqlistcopy .txt  
Output Set: N:\CRF4\04212003\I988292A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; Xaa Pos.7,165

Seq#:2; Xaa Pos.7,165

Seq#:5; N Pos. 829,868,1018,1037,1041

Seq#:6; N Pos. 470,485

Seq#:8; N Pos. 775,817,844,846,856,857,859,861,863,871,872

Seq#:21; N Pos. 70,366,388,389,399

Seq#:22; N Pos. 361,382,385,446,447,469,471,536,581,584,586

## VERIFICATION SUMMARY

DATE: 04/21/2003

PATENT APPLICATION: US/09/988,292A

TIME: 15:30:58

Input Set : A:\pfl60d2seqlistcopy .txt

Output Set: N:\CRF4\04212003\I988292A.raw

L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:65 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:48  
M:341 Repeated in SeqNo=1  
L:124 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2 after pos.:0  
M:341 Repeated in SeqNo=2